Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 05, May -2018

# FAST PHRASE SEARCH FOR ENCRYPTED CLOUD STORAGE

Priyanka Raut[1st],Nikhil Lohakare[2nd],Sheetal Kolekar[3rd],Prof.Sneha Pisey[4th]

*Modern Education Society's College of Engineering, Pune*

**Abstrac**t:- —*Cloud computing has generated abundant interest within the analysis community in recent years for its several benefits, however has additionally raise security and privacy considerations. The storage and access of confidential documents are identified together of the central issues within the space. especially, several researchers investigated solutions to go looking over encrypted documents keep on remote cloud servers. whereas several schemes are planned to perform conjunctive keyword search, less attention has been noted on additional specialized looking out techniques. During this paper, we have a tendency to gift a phrase search technique supported Bloom filters that's significantly quicker than existing solutions, with similar or higher storage and communication price. In this project, at the time of file uploading on cloud we check file deduplication. We store only unique files on cloud. Using MD5 Algorithm We check file deduplication. File deduplication checking is used for cloud storage management. Our technique uses a series of n-gram filters to support the practicality. The theme exhibits a trade-off between storage and false positive rate, and is filmable to defend against inclusion-relation attacks. A style approach supported Associate in Nursing application's target false positive rate is additionally represented. Secure data deduplication can significantly reduce the communication and storage overheads in cloud storage services, and has potential applications in our big data-driven society.*

***Keywords:-*** *Conjunctive keyword search, file deduplication, Phrase search, Privacy, Security, Encryption*

## INTRODUCTION

CLOUD computing has emerged as a disruptive trend in both IT industries and research communities recently, its salient characteristics like high scalability and pay-as you-go fashion have enabled cloud consumers to purchase the powerful computing resources as services according to their actual requirements, such that cloud users have no longer need to worry about the wasting on computing resources and the complexity on hardware platform management. Nowadays, more and more companies and individuals from a large number of big data application shave outsource their data and deploy their services into cloud servers for easy data management, efficient data mining and query processing tasks. Data encryption has been widely used for data privacy preservation in data sharing scenarios, it refers to mathematical calculation and algorithmic scheme that transform plaintext into cipher text, which is a non-readable form to Secure data deduplication can significantly reduce the communication and storage overheads in cloud storage services,
And has potential applications in our big data-driven society. Unauthorized parties. A variety of data encryption models have been proposed and they are used to encrypt the data before outsourcing to the cloud servers. However, applying these approaches for data encryption usually cause tremendous cost in terms of data utility, which makes traditional data processing methods that are designed for plain text data no longer work well over encrypted data . Secure data deduplication can significantly reduce the communication and storage overheads in cloud storage services, and has potential applications in our big data-driven society.
Data encryption has been wide used for knowledge privacy preservation in knowledge sharing situations, it refers to mathematical calculation and algorithmic theme that remodel plaintext into cipher text, that may be a non-readable type to unauthorized parties. a spread of knowledge secret writing models have been planned and that they square measure accustomed write in code the data before outsourcing to the cloud servers. However, applying these approaches for encryption typically cause tremendous price in terms of knowledge utility, that makes traditional processing strategies that square measure designed for plaintext knowledge now not work spill encrypted knowledge. Data encryption has been wide used for knowledge privacy preservation in knowledge sharing situations, it refers to mathematical calculation and algorithmic theme that remodel plaintext into cipher text, that may be a non-readable type to unauthorized parties. a spread of knowledge secret writing models have been planned and that they square measure accustomed write in code the data before outsourcing to the cloud servers. However, applying these approaches for encryption typically cause tremendous price in terms of knowledge utility, that makes traditional processing strategies that square measure designed for plaintext knowledge now not work spill encrypted knowledge.

## COMMUNICATION FRAMEWORK

We'll describe our keyword search framework using two parties: The data owner and an untreated cloud server. Our algorithms can easily be adapted to the scenario of an organization wishing to setup a cloud server for its employees by implementing a proxy server in place of the data owner and having the employees/users authenticate to the proxy server. A standard keyword search protocol is shown in figure. During setup, the data owner generates the required encryption

keys for hashing and encryption operations. Then, all documents in the database are parsed for keywords. Bloom filters tied to hashed keywords and grams are attached. The documents are then symmetrically encrypted and uploaded to the cloud server. To add files to the database, the data owner parses the files as in setup and uploads them with Bloom filters attached to the cloud server. To remove a file from the data , the data owner simply sends the request to the cloud server, who removes the file along with the attached Bloom filters.

**Security**

In terms of security, we assume a semi-honest cloud server, which is interested in learning about stored data but will follow our keyword search protocol as described and will not modify or misrepresent any data in order to gain an advantage. Two of the main security issues regarding keyword searches are the privacy of the document sets and the privacy of the queried keywords. Briefly, a secure keyword search protocol should prevent the cloud server from obtaining non-negligible amount of information on the stored documents or the keywords in the query requests. Note that, in our target application, users are employees of the data owner's organization and are authorized to search for any documents in the data set. Should an application requires that users be restricted from accessing certain files, an access control system such as would be required to verify the matched results and returned only those which the user has the required credential to access. Our basic scheme in section achieves these goals under the assumption that the cloud has no prior knowledge on the stored data. Should the cloud provider has significant statistical knowledge on the stored data, such as the distribution of the keywords, it may be able to infer partial knowledge on its content. Under the security model where the cloud provider has some knowledge over the distribution of keywords or queries on the stored data, we describe modifications to the basic scheme which would offer protection against statistical attacks in section 4.6 and inclusion-relation

## LITERATURE SURVEY

**Paper1**:A Secure and Dynamic Multi Keyword Ranked Search Scheme over encrypted.

The major aim of this paper is to resolve the matter of multi-keyword hierarchical search over encrypted cloud knowledge (MRSE) at the time of protective actual technique wise privacy within the cloud computing construct. knowledge holders area unit inspired to source their tough knowledge management systems from native sites to the business public cloud for big flexibility and monetary savings. but for protecting knowledge privacy, sensitive knowledge got to be encrypted before outsourcing, which performs ancient knowledge utilization supported plaintext keyword search. As a result, permitting Associate in Nursing encrypted cloud knowledge search service is of supreme significance. visible of the massive range of information users and documents within the cloud, it's essential to allow many keywords within the search demand and come back documents within the order of their acceptable to those keywords. Similar mechanism on searchable cryptography makes centre on single keyword search or Boolean keyword search, and infrequently type the search results. within the middle of various multi-keyword linguistics, deciding the well-organized similarity live of coordinate matching, it means as several matches as doable, to capture the suitable knowledge documents to the search question. notably, we consider dot product similarity i.e., the number of question keywords shows in a document, to quantitatively estimate such match live that document to the search question. Through the index construction, each document is connected with a binary vector as a sub index wherever every bit characterize whether or not matching keyword is contained within the document.

**Paper2**: Privacy-Preserving Multi-keyword Ranked Search Over Encrypted Cloud Data, The innovation in cloud computing has encouraged the data owners to outsource their data managing system from local sites to profitable public cloud for excessive flexibility and profitable savings. But people can like full benefit of cloud computing, if we are able to report very real secrecy and security concerns that come with loading sensitive personal information. Allowing an encrypted cloud data search facility is of great significance. In view of the huge number of data users, documents in the cloud, it is important for the search facility to agree multi keywords query and arrange for result compare.

**-Paper3**: Secure Indexes.

A secure index is a data structure that allows a queried with a trapdoor for a word x to test in O(1) time only if the index contains x; The index reveals no information about its contents without valid trapdoors, and trapdoors can only be generated with a secret key. Secure indexes are a natural extension of the problem of constructing data structures with privacy guarantees such as those provided by oblivious and history independent data structures. In this paper, we formally define a secure index and formulate a security model for indexes known as semantic security against adaptive chosen keyword attack (ind-cka). We also develop an efficient indicia secure index construction called z-index using pseudo-random functions and Bloom filters, and show how to use z-idx to implement searches on encrypted data. This search scheme is the most efficient encrypted data search scheme currently known; It provides O(1) search time per document, and handles compressed data, variable length words, and Boolean and certain regular expression queries. The techniques developed in this paper can also be used to build encrypted searchable audit logs, private database query schemes, accumulated hashing schemes, and secure set membership tests.

**Existing System:**

In existing system , Their scheme uses public key encryption to allow keywords to be searchable without revealing data content but investigated the problem for searching over encrypted audit logs. Many of the early works focused on single

keyword searches. Recently, researchers have proposed solutions on conjunctive keyword search, which involves multiple keywords. Other interesting problems, such as the ranking of search results and searching with keywords that might contain errors termed fuzzy keyword search, have also been considered.

Existing System Disadvantages:

- Single keyword search is not smart enough to support advanced queries .
- Boolean search is unrealistic since it causes high communication cost

**OBJECTIVE**

1. Big data encryption against privacy break.

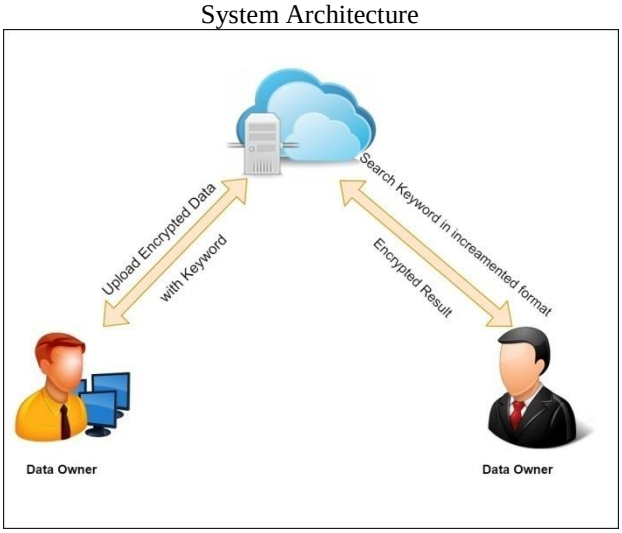
2. Improve the capability of defending the privacy break.
3. Improve scalability and the time efficiency of query processing.

**Proposed System:**

we present a phrase search scheme which achieves a much faster response time than existing solutions. The scheme is also scalable, where documents can easily be removed and added to the collection. We also describe modifications to the scheme to lower storage cost at a small cost in response time and to defend against cloud providers with statistical knowledge on stored data. We begin by presenting the communication framework and various backgrounds including related works . Although phrase searches are processed independently using our technique, they are typically a specialized function in a keyword search scheme, where the primary function is to provide conjunctive keyword searches. We describe both basic conjunctive algorithm and phrase search algorithm. In this project, at the time of file uploading on cloud we check file deduplication. We store only unique files on cloud. Using MD5 Algorithm We check file deduplication. File deduplication checking is used for cloud storage management.

Trapdoor: In Cloud Computing we use this method to encrypt our data with various random possibilities to make our data more secure on cloud. And the same procedure is follow to decrypt our data in order to reuse our file or any data.

System Architecture

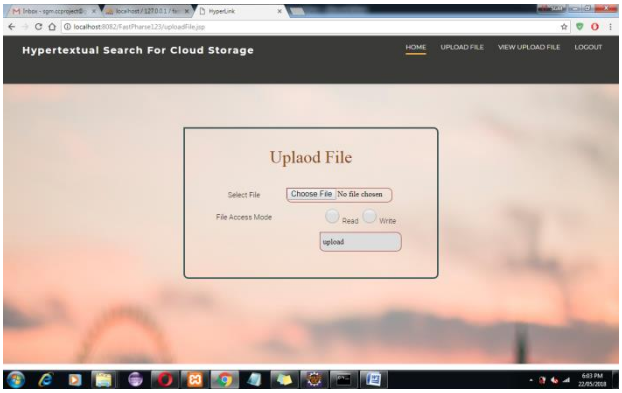



**CONCLUSION AND FUTURE SCOPE**

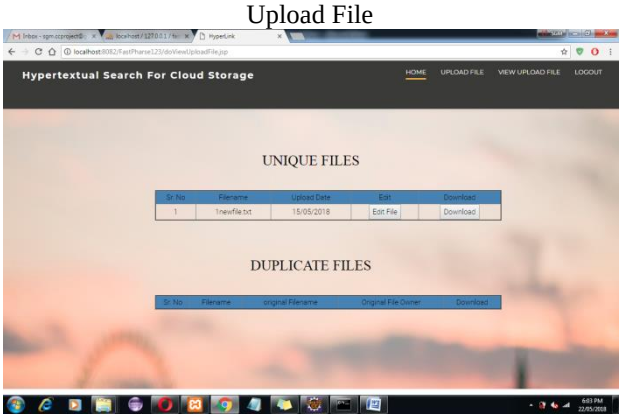
We best owed a phrase search theme supported Bloom alter that's considerably quicker than existing approaches, requiring solely one spherical of communication and Bloom filter verifications. the answer addresses the high process price noted in by reformulating phrase search as n-gram verification instead of a location search or a sequent chain verification. not like our schemes take into account solely the existence of a phrase, omitting any info of its location. not like our schemes don't need sequent verification, is parallelizable and incorporates a sensible storage demand. In this project, at the time of file uploading on cloud we check file deduplication. We store only unique files on cloud. Using MD5 Algorithm We check file deduplication. File deduplication checking is used for cloud storage management. Our approach is additionally initial the primary to effectively permit phrase search to run severally while not first performing arts a conjunctive keyword search to spot candidate documents. The technique of constructing a Bloom filter index introduced in section allows quick verification of Bloom filters within the same manner as classification.

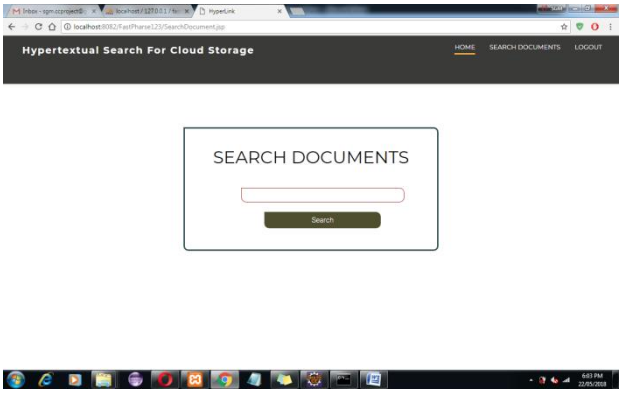
RESULT


Upload File

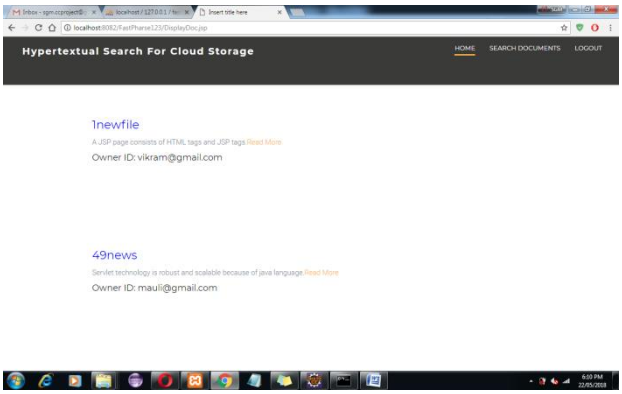



View Files

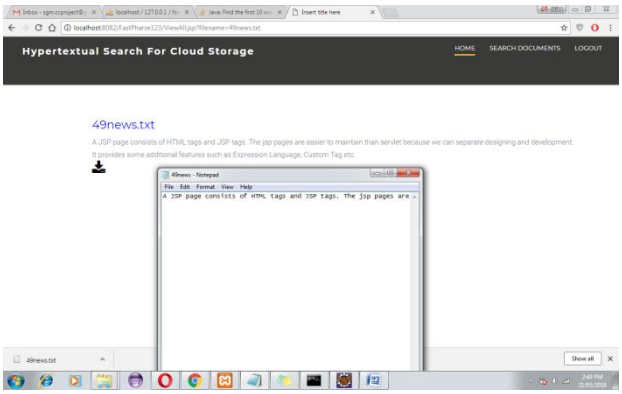



Search FIle




View File




Result

## REFERENCES

[1] R. Carmela, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions,"in Proceedings of the 13th ACM Conference on Computer and Communications Security. ACM, 2006, pp. 79–88.

[2] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. SP 2000. Proceedings. 2000 IEEE Symposium on, 2000, pp. 44–55.

[3] E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, p. 216, 2003.

[4] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014.

[5] Z. Xia, X. Wang, X. Sun, and Q. Wang, "A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 2, pp. 340–352, 2016

[6] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. SP 2000.Proceedings. 2000 IEEE Symposium on, 2000, pp. 44–55.

[7] E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive,vol. 2003, p. 216, 2003

[8] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Applied Cryptography and Network Security. Springer, 2005, pp. 442–455

[9] Y. H. Hwang and P. J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in Pairing-Based Cryptography–Pairing. Springer, 2007, pp. 2–22.

[10] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Applied Cryptography and Network Security. Springer, 2004, pp. 31–4